اونيۏرسيتي تيكنولوجي مارا
UNIVERSITI TEKNOLOGI MARA

University Publication Centre (UPENA)

ESTEEM

Academic Journal UiTM Pulau Pinang

Volume 6, Number 1 June 2010 ISSN 1675-7939

Esteem Academic Journal is jointly published by the Universiti Teknologi MARA, Pulau Pinang and University Publication Centre (UPENA), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

ESTEEM

Academic Journal UiTM Pulau Pinang

Volume 6, Number 1 June 2010 ISSN 1675-7939

Foreword

It is indeed a proud moment for the University Publication Centre (UPENA) of UiTM Pulau Pinang for having realised the publication of the sixth volume of the Esteem Academic Journal UiTM Pulau Pinang. In fact, it is the undivided support and all-round commitment from all those who were directly and indirectly involved in this project that was the pivotal factor for this success.

On behalf of UPENA UiTMPP, I would like to, first and foremost, express my sincerest gratitude to Associate Professor Mohd Zaki Abdullah, Director of UiTM Pulau Pinang, Associate Professor Dr Mohamad Abdullah Hemdi, Deputy Director of Academic Affairs and Associate Professor Ir. Damanhuri Jamalludin, Deputy Director of Research, Industry Linkages, Development & Maintenance for their unwavering support and being such a driving force towards this successful endeavour.

Not to be forgotten also is the service rendered by the distinguished panel of external reviewers for their constructive comments and criticisms in ensuring that the papers published in this issue would be of the highest quality. Similarly, the panel of language editors who had worked tirelessly towards ensuring that the papers published were linguistically perfect. To both these groups, UPENA is in awe of your efforts and salutes you!

UPENA is also impressed with the nature of papers submitted for publication. While this issue comprises all engineering based articles, it covers a wide array of sub-engineering disciplines. Kudos to these writers! UPENA sincerely appreciates their efforts and hopes more of our staff will follow in their footsteps.

Finally, research and publication are integral parts of an academic's life at any institution. Apart from being an institutional requirement, it is also essential for our own continuous self-development and knowledge expansion. To this effect, UPENA hopes to play a significant role by providing the platform upon which our staff can realise their dream. So, it is our hope at UPENA UiTMPP that lecturers will take up the challenge and start to publish more vigorously from now on.

Rasaya Marimuthu
Chief Editor
ESTEEM Vol. 6, No. 1, 2010
(Engineering)

Esteem Academic Journal *Vol. 6, No. 1, 53-72, 2010*

Using Intraclass Correlation Coefficient and Bartlett Test Statistic to Identify Soil Layer Boundaries

Lim Jit Kheng
Faculty of Civil Engineering
Universiti Teknologi MARA (UiTM), Malaysia
Email: limjitkheng@yahoo.com

Ng Set Foong
Department of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM), Malaysia

Mohamad Razip Selamat
School of Civil Engineering
Universiti Sains Malaysia, Malaysia

Eric Goh Kok Hoe
School of Materials and Mineral Resources Engineering
Universiti Sains Malaysia, Malaysia

ABSTRACT

In this study, the performance of two statistical methods, namely intraclass correlation coefficient (RI) and Bartlett test statistic in conjunction with various suggested window widths are investigated to identify soil layer boundaries. The study is done using three fairly different CPT soundings obtained from the database of National Geotechnical Experimental Sites. The identification of layer boundaries and demarcating the soil profile into homogeneous layers is very important in geotechnical engineering. From this study, RI appears to be a more powerful, robust and persistent tool and the corresponding suitable window width was proven as a function of average

ISSN 1675-7939

distance between boundaries which could be determined from autocorrelation analysis. Furthermore, a simple approximate method is also proposed in this study to estimate the suitable window width using the concept of average distance between 'mean-crossings'. The approach was exploited and substantiated as a simple, quick and accurate estimator in making the first approximation on suitable window width for boundary identification exercise.

Keywords: *Soils; Boundary layers; Homogeneity; Statistics; Stationary processes.*

Introduction

A major uncertainty in geotechnical engineering is the inherent spatial variability of soil properties. The importance of recognizing uncertainties and taking them into account in geotechnical design have been propagated by numerous leaders since 1960s (Casagrande, 1965; Peck, 1969; Wu, 1974; Leonards, 1982; Tang, 1984; Morgenstern, 1995; Whitman, 2000; Christian, 2004). Probability theory and statistics provide a formal, scientific and quantitative basis in assessing risk and uncertainties and have been sprouted in geotechnical engineering research in recent years. In line with the development, characterization of soil properties has been advanced onto the functions of the deterministic mean and its stochastic characters, comprises of coefficient of variation and scale of fluctuation in modeling the inherent soil variability as a random field (Vanmarcke, 1977; DeGroot and Baecher, 1993; Jaksa et al., 1997; Phoon and Kulhawy, 1999; Fenton, 1999).

Compliance to stationarity or statistical homogeneity criterion is imperative in any soil data analyses. A random function which used to modeled the variability of soil is considered stationary, or weakly stationary if (Brockwell and Davis, 2002): (1) the mean of the function is constant with distance, i.e. no trend in the data, (2) the variance of the function is constant with distance, i.e. homoscedastic, (3) there are no seasonal variations, (4) there are no apparent changes in behaviors, and (5) there are no outlying observations. In other words, a stationary series is essentially a function of their separation distance, rather than their absolute locations.

In geotechnical characterization undertaking, the first step usually involves demarcating the soil profile into layers or sections which are homogeneous so that the result of subsequent analysis is not biased. A homogeneous layer comprises uniform soil material that has undergone similar geologic history and possesses with certain distinctive behaviors. The identification of boundaries and thus demarcation process are often much more complicated than one expected when dealing with this highly variable complex natural material. The variability exists not only from site to site and stratum to stratum, but even within apparently homogeneous deposits at a single site (Baecher and Christian, 2003).

Thus, it would be rather useful to supplement the existing procedures with a quantitative systematic approach. The conventional method, which is based on visual observation, gives less accuracy and substantial subjectivity to the identification of actual boundary of soil. Existing statistical tools are not widely explored, well-calibrated and properly defined, thus generally result in unsatisfactory outcome. This paper intends to resolve the above problem for better characterization of soil properties. Reported useful statistical tools would be compared in terms of their effectiveness and the existing procedures are revamped for further improvement.

Cone Penetration Test (CPT) is widely used in soil characterization in view of its ability to provide almost continuous profile, widely correlated and highly repeatable (Robertson, 1986; NCHRP, 2007). In this study, the CPT soundings, in particular the cone tip resistance, q_c, were used for detail illustration. The data were selected on the basis of their spacing, extensiveness and difference between each other which best yield the thorough examination on the performance of statistical tools in soil boundary demarcation.

Statistical Approaches

Classical and advance statistical approaches for testing similarity or dissimilarity of univariate or even multivariate records are believed to be substantial. Some of the established analytical tools have potential to be applied in the field of geotechnical engineering with modification to suit the nature of geotechnical parameters. In this paper, two statistical methods which are relatively common and simple in identifying the soil layer boundaries are presented.

Intraclass Correlation Coefficient

Intraclass Correlation Coefficient (RI) was reported by Campanella and Wickreseminghe (1991) as a useful statistical method for detecting soil layer boundaries using CPT soundings. For identification of layer boundaries, a moving window width, W_d, is first specified and the window is divided into two segments. The RI profile is then generated by moving two contiguous segments over a measurement profile and the computed index is plotted corresponding to the midpoint of the window. RI will always lie between zero and unity and a relatively high value of RI is likely to indicate the presence of a layer boundary.

The RI together with its pooled combined variance (s_w^2) and the between class variance (s_b^2) are defined as

$$\mathrm{RI} = \frac{s_b^2}{s_b^2 + s_w^2} \tag{1a}$$

$$s_w^2 = \frac{n_1}{n_1 + n_2 - 1} s_1^2 + \frac{n_2}{n_1 + n_2 - 1} s_2^2 \tag{1b}$$

$$s_b^2 = s^2 = \frac{1}{n_1 + n_2 - 1} \sum_{i=1}^{n_1+n_2} (x_i - \bar{x})^2 \tag{1c}$$

where n_1 and n_2 are the sample size of two equal segments, above and below the middle line of the window, s_1^2 and s_2^2 being the variances of the sample for the two segments, $\bar{x}$ and s^2 are the sample mean and variance within the designated window. The equation can also be written as follow (Zhang and Tumay, 1996) for the two segments with equal sample size of m and their sample mean of $\bar{x}_1$ and $\bar{x}_2$, respectively.

$$\mathrm{RI} = \frac{1}{1 + \dfrac{1}{\dfrac{m-1}{m} + \dfrac{(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 + s_2^2)}}} \tag{1d}$$

Judging whether an index value is high enough to indicate a boundary in a relative sense by visual observation is fairly subjective and would result in inconsistency. Zhang and Tumay (1996) suggested that the peak value of RI which is equal to or larger than 0.7 can be empirically determined as the boundary line. Hegazy et al. (1996) proposed the critical value as the (mean + 1.65 standard deviation) with a level of significance of 5%. However, Phoon et al. (2003) commented that the above critical values are not depending on the underlying correlation structure of the profile.

Bartlett Test

The classical Bartlett test is used to test the equality of multiple sample variances for independent data sets. The test is sensitive to departures from normality as well as heteroscedasticity, thus is rather nonrobust. However, when it is applicable, it is more powerful than various other tests. Phoon et al. (2003) adopted this statistical method in identifying soil layer boundary by moving a sampling window over the soil profile to generate a continuous Bartlett test statistic profile. For the case of two sample variances, the Bartlett test statistic (Bstat) is reduced to

$$\text{Bstat} = \frac{4.605(m-1)^2}{2m-1}\left[2\log s^2 - (\log s_1^{\,2} + \log s_2^{\,2})\right] \qquad (2a)$$

or

$$\text{Bstat} = \frac{2(m-1)^2}{2m-1}\left[2\ln s^2 - (\ln s_1^{\,2} + \ln s_2^{\,2})\right] \qquad (2b)$$

where m is the number of data points in each two equal segments of the sampling window, and $s_1^{\,2}$, $s_2^{\,2}$ and s^2 being the variances of the sample for the first segment, second segment and their total, respectively. Note that where the $s_1^{\,2}$ and $s_2^{\,2}$ are equal, the Bstat is zero. Where they are largely different, a relatively high value of Bstat is computed, it is likely to indicate the presence of a layer boundary.

To reject the null hypothesis of stationarity, Phoon et al. (2003) commented that the rejection criteria for classical Bartlett test is not applicable to correlated data, thus proposed the modified Bartlett test statistic which claimed to be more discriminating than other traditional classical test statistics as it incorporates the correlation structure in the underlying data. Critical values corresponding to 5% level of significance for various common autocorrelation models were generated from simulated correlated sample functions as the appropriate rejection criteria.

Window Width

As all these statistical methods incorporate with the concept of moving windows, the width of the sampling window becomes an important parameter which could have substantial influence on the result of analysis. Generally, too narrow a window will result in undesirable effect of high noise level with too many peaks appear. On the other hand, too wide a window will over-smoothen the statistics till missing out the possible boundaries due to excessive perturbation region.

Webster (1973) proposed a method to determine the boundaries on transects automatically and has found that the suitable width for the calculation window is approximately two thirds of the expected distance between boundaries where the spacing between boundaries does not differ widely. The expected distance or average spacing between boundaries could be determined from an autocorrelation analysis. The technique is said reasonably sensitive but found little affected by window width.

Campanella and Wickremesinghe (1991) elaborated in detail the statistical methods for determination of window width and recommended that it is rather to adopt an incorrect narrower window width than wider window width to avoid missing out the possible layer boundaries. Two case studies, namely McDonald farm Site and Haney Site have been illustrated and the window widths selected were 1.5 m and 2.0 m, respectively. To the other extent, window width that less than 1.0 m should not be selected due to normal distribution restriction on the samples (Wickremesinghe, 1989).

Zhang and Tumay (1996) based on the finding of previous research that the standard 10 cm^2 electric cones may require minimum stiff layer thickness of 36 cm to 72 cm to ensure full tip resistance and concluded that the value of the window width could be conservatively taken to be 150 cm or 75 cm for half of the window. Nevertheless, they reported that primary layering usually does not provide satisfactory results due to uneven soil layers. The big difference of layer thicknesses will result in too many layers in thick zone and too little in the thin causing a bias in making a judgment.

Cafaro and Cherubini (2002) used the same procedure as proposed by Webster (1973) in analyzing a stiff overconsolidated clay at a test site in Taranto, Southern Italy and obtained a fairly wide window width of 6.8 m for q_c, f_s and I_c profiles. The width was reduced to 4.8 m and the variation in the RI profile was found negligible on both the position (depth) and the value of the peaks. The geostatistical boundary was found not always correspond well to the geolithological boundary thus suggested the possible offset to be accounted for.

Kulatilake and Um (2003) introduced a new procedure to detect statistical homogeneous layers in a soil profile. In examining the cone tip resistance data for the clay site at Texas A&M University, a window width of 0.4 m which contains 10 data points in each section has been used. Due to the short section adopted, four possible combinations for the mean soil property (either constant or with linear trend) were considered. The distance between the lower and upper sections was calculated and subsequently generated along the depth for evaluation of the statistical homogeneity at different levels.

Phoon et al. (2003) adopted the lower limit of permissible window width, that is 1.0 m in generating the RI and Bstat profiles. Both profiles manage to capture the primary layer boundaries in consistent with visual inspection, of which the Bstat peaks were much more prominent. Considerable noises were observed and 3 fault boundaries were identified (no obvious soil boundaries can be seen in the q_t record) in the RI profile when comparing to the critical value of 0.7.

Autocorrelation Analysis

The Webster's intuition that the suitable window width should be equal to or somewhat less than the average distance between boundaries has led to the exploitation of autocorrelation analysis (Webster, 1973). The result of his study showed that the optimize width is around two thirds of the expected distance although larger width up to the full expected distance could still be useful (main peaks appeared in the same positions but with different relative heights) for those area with marked changes. The autocorrelation coefficient at lag k was expressed as

$$r_k = \left[\frac{n}{n-k}\right]\left[\frac{\sum_{t=1}^{n-k}\left(u_t u_{t+k}\right)}{\sum_{t=1}^{n} u_t^2}\right]$$

where n is the number of sampling points in the series, k is the lag and u is the deviation from the series mean at the t^{th} (or $t + k^{th}$) point.

In the correlogram, i.e. the plot of autocorrelation coefficient, r_k against lag, k, the autocorrelation coefficient will decrease more or less steadily with increasing lag distance from around 1 to some minimum value near zero and fluctuate thereafter. The lag distance over which this decay occurs can be taken as the average distance between boundaries which could be used as guidance in selection of suitable window width.

Numerical Experiments

For this study, established database for National Geotechnical Experimental Sites (NGES) (http://www.unh.edu/nges/) funded by the Federal Highway Administration (FHWA) and National Science Foundation (NSF) of America were explored. The performance and usefulness of several approaches (in terms of window width and the statistical tool) that have been used by other geotechnical researchers in identifying layer boundaries were thoroughly examined. Typical CPT profiles from three sites representing different parts of North America were selected. The sites are Treasure Island Naval Station in the San Francisco Bay area (CATIFS), University of Massachusetts, Amherst campus (MAUMASSA), and Northwestern University Lake Fill Site in Evanston (ILNWULAK). These sites have been classified as Level I and II sites that are most closely fit the combined criteria of research areas as of significant national importance. These CPT soundings were closely spaced, extensive and fairly different between each other which best suit for this examination.

Examination of Existing Approaches

Two statistical methods, namely intraclass correlation coefficient (RI) and classical Bartlett test statistic (Bstat) which reported to be useful in geotechnical literatures are presented here. Since the methods are used in conjunction with moving window averaging concept, the sensitivity of the window width in generating the most optimize profiles which would discriminate the 'true' boundaries is of great concern. Several criteria in determining suitable window width deduced from previous researchers' works are incorporated in this study as follows:

i. Two thirds of the average distance between boundaries determined by using autocorrelation analysis (Webster, 1973; Campanella and Wickremesinghe, 1991),
ii. The minimum thickness for full tip resistance, i.e. 1.5 m (Zhang and Tumay, 1996), and
iii. The lower permissible limit of window width, i.e. 1.0 m (Wickremesinghe, 1989; Phoon et al., 2003).

In reality, perfect result is almost not possible as the actual soil data could be really erratic. However, the analytical approach (combination of a statistical tool with an optimize window width) could be deemed satisfactory from at least two practical aspects. The approach should avoid missing out possible prominent layer boundaries and at the same time, able to capture as many major boundaries as possible at one time. And it is evident from past literatures that the boundaries indicated by these statistical tools are often slightly offset probably due to variation in the upper and lower segments as moving the sampling window over the soil profile. Therefore, note that the tools could serve as a useful indicator, but yet final adjustment and decision have to be made with regards to the original profile, geological background and not forgotten engineering judgment.

The results of analysis for CATIFS, MAUMASSA and ILNWULAK sites are presented in Figure 1, Figure 2 and Figure 3, respectively. For each set of results, 3 different window widths as delineated above (i, ii and iii) together with the full expected distance between boundaries (Webster, 1973) have been incorporated with both RI and Bstat and presented side-by-side for comparison. Figure 4 shows the plot of autocorrelation coefficient against lag distance for cone penetration resistance at CATIFS site.

From the cone tip resistance profile in Figure 1 (CATIFS site), it can be observed through visual examination that there is soil boundary for soil from 7.0 m to 9.0 m. The RI profile manages to capture these peaks (with RI generally more than 0.7) at both 7.0 m and 9.0 m locations, and another one at approximately 2.1 m. The above three main peaks were found persist for all the tested widths of 1.0 m, 1.5 m, 1.9 m and 2.8 m (Figure 1(a)-1(d)). However, there are more peaks (exceeding the empirical value of 0.7) for window widths of 1.0 m and 1.5 m (Figure 1(a) and 1(b)). Thus, inference could be made from the results that any width which falls within the range of 1.9 m and 2.8 m are suitable in this case.

On the other hand, due to inapplicable of classical rejection criteria (Phoon et al., 2003), Bstat profile does not have any valid guidance in making decision whether the peaks are sufficiently high to reject the null hypothesis of weak stationarity in the preliminary stage. Therefore, only the apparently relative discriminating peaks were highlighted as potential boundaries in this study. A main peak at approximately 7.3 m was recognized in Bstat profiles with window widths of 1.0 m and 1.5 m (Figure 1(a) and 1(b)). However, as the window width increased to 1.9 m and subsequently 2.8 m (Figure 1(c) and 1(d)), two main peaks were appeared at approximately 7.3 m and 8.6 m with fairly different profile

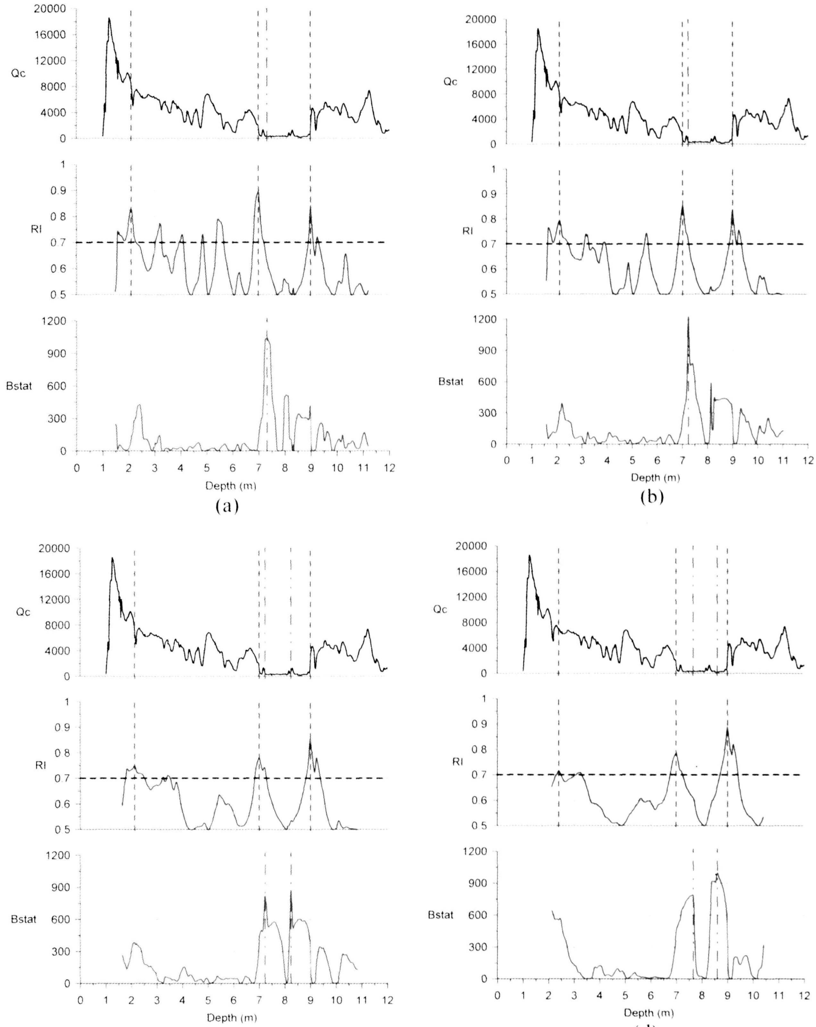


Figure 1: Identification of Soil Layer Boundaries using RI and Bstat of Varying W_d for CATIFS site. (a) $W_d = 1.0$ m, (b) $W_d = 1.5$ m, (c) $W_d = 1.9$ m, and (d) $W_d = 2.8$ m

among them. Other moderate peaks (e.g. around 2.1 m and 9.4 m) can hardly be concluded as potential boundaries. In general, Bartlett test method failed to adequately identify the major prominent boundaries with their respective reasonable accurate locations in this case.

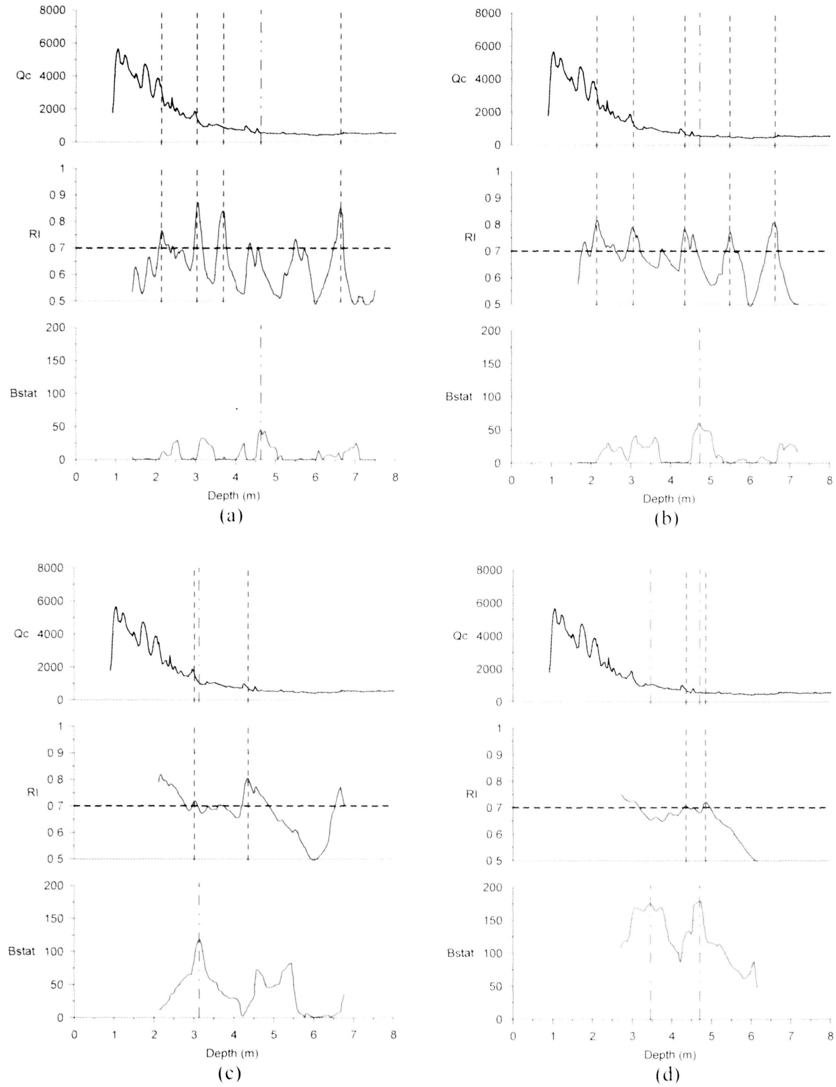


Figure 2: Identification of Soil Layer Boundaries using RI and Bstat of Varying W_d for MAUMASSA site. (a) W_d = 1.0 m, (b) W_d = 1.5 m, (c) W_d = 2.4 m, and (d) W_d = 3.6 m

The limitation of both approaches on missing out the information at both ends of the profile is readily observed. The apparent boundary of cone tip resistance at 1.26 m for instance is basically out of the coverage area of the generated output as the computed index is plotted against the

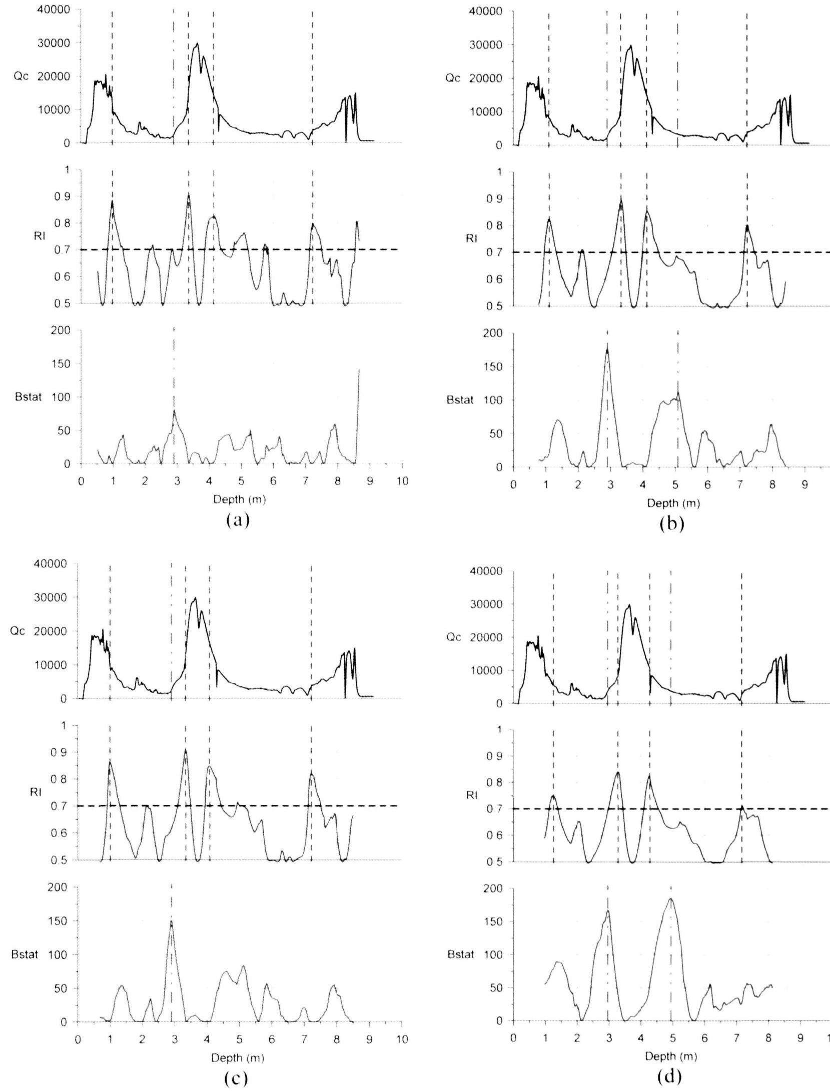


Figure 3: Identification of Soil Layer Boundaries using RI and Bstat of Varying W_d for ILNWULAK Site

midpoint of the moving window. In addition, the identification of a potential boundary around the depth of 2.1 m using RI profile indicated that the tool able to detect a considerable sharp change along the profile which suggested two quasi-linear portions to be divided.

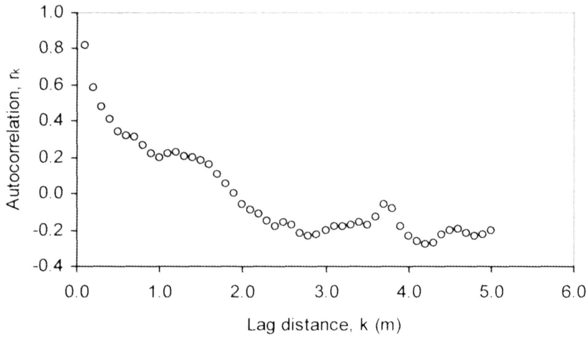


Figure 4: Empirical Autocorrelation Function for Cone Penetration Resistance at CATIFS Site

MAUMASSA site (Figure 2) was the second case study where the cone tip resistance profile exhibits apparent heteroscedasticity characteristic with gradual change of gradient around the potential boundary. As shown in Figure 2(a) and 2(b), the generated RI profiles for window widths of 1.0 m and 1.5 m are basically a noise inferring that the windows are too narrow. As the window width increased to 2.4 m (Figure 2(c)), which is approximately two thirds of the average distance between boundaries, the 'true' main peaks appeared, one at approximate depth of 4.4 m and another one at 3.0 m. Generated profile at both ends tends to be less reliable as shown by two erroneous peaks thus should not be considered. Further increment on the width to 3.6 m (Figure 2(d)) has over-smoothen the profile and tends to hidden the boundary at 4.4 m (marginally above 0.7) and causing undesired peak at around 4.8 m. Also, substantial information at both ends of the profile was sacrificed where potential boundary there might be missed out. In this case, window width between 2.4 m and 3.6 m (preferably closer to 2.4 m than 3.6 m) can be considered as the suitable width as RI is concerned.

Bstat profile suffered from the same problem as discussed in previous case where difficulty was faced while interpreting the result to decide whether or not the peaks need to be considered as boundaries. For window widths of 1.0 m and 1.5 m (Figure 2(a) and 2(b)), Bstat profile experienced the same noisy problem as RI but consistently showing a potential boundary at approximate depth of 4.7 m if the most prominent peak to be valid. However, the main peak changed to 3.1 m for window width of 2.4 m (Figure 2(c)) and subsequently a vague wide peak around 3.0 m to 3.7 m with another one at 4.7 m when the window widths increased to 3.6 m (Figure 2(d)). From the result above, Bstat seems to be less robust than RI in identifying potential boundaries.

Results of the third case study of ILNWULAK site are presented in Figure 3. The cone tip resistance profile seems to exhibit higher resistance values at both ends, i.e. depths before 1.0 m and after 7.0 m, and it is observed through visual examination that an interbedded heterogeneous layer at approximately 3.3 m to 4.3 m. RI profiles for all the tested widths of 1.0 m, 1.5 m, 1.3 m and 1.9 m as presented in Figure 3(a) to 3(d) show a very good agreement where the four expected boundaries were managed to detect. The main peaks persist as the window width changes and more noises can be noticed at smaller widths particularly for window width of 1.0 m. Similar inference could still be reasonable drawn where the suitable width for this boundaries demarcation exercise ranges from 1.3 m to 1.9 m (preferably closer to 1.9 m), obtained from the autocorrelation analysis.

For Bartlett test, the Bstat profile appears to be quite sensitive to the changes of window width. Initial trial using window width of 1.0 m (Figure 3(a)) resulted in a fairly prominent peak located at approximate depth of 2.9 m. The same peak becomes more apparent as the width increases to 1.3 m and 1.5 m (Figure 3(c) and 3(b)) with another pretty vague peak noticed around 5.0 m. Subsequently, further increment of the window width to 1.9 m (Figure 3(d)) has resulted in two prominent peaks detected at 3.0 m and 5.0 m. In this case, RI has been found to be able to detect most of the major boundaries at one time reasonably accurately compared to Bstat.

In general, both statistical tools can be utilized in identifying layer boundaries with their own advantages and disadvantages. RI appears to be a more powerful tool as it can capture most of the prominent major boundaries at one time fairly accurately. It is also relatively more robust which could persistently detect the main peaks at the same positions even with window widths that are fairly different from the optimize configuration. Webster's suggestion (1973) to determine the suitable window width as a function of average distance between boundaries using autocorrelation analysis was validated. The difference for profiles generated using smaller window widths is that many undesired peaks or noises may appear, whereas larger widths tend to hidden the necessary boundaries. The empirical criterion of 0.7 (Zhang and Tumay, 1996) as a guidance in deciding whether a peak is significant enough to be considered as a valid boundary is very useful. The criterion was found performing pretty well in most circumstances as illustrated through various distinctive case studies in this paper.

The main problem for Bstat is that no applicable rejection criterion can be referred to in judging the existence or nonexistence of a boundary (Phoon et al., 2003). The proposed modified Bartlett test statistics by Phoon et al. (2003) is pretty complex as it involves autocorrelation analysis and scale of fluctuation in deriving the critical values. It tends to be a useful approach for rigorous stationarity check rather than a simple method in identifying potential homogeneous layers at the preliminary stage. Bstat is more sensitive to the effect of outliers and departures from normality. The acceptable window width for Bstat can be pretty smaller than that required for RI, i.e. at around 1.0 m regardless of the layer thickness to yield the prominent peak representing soil layer boundary. In addition, Bstat can only identify limited layer boundaries (usually those most prominent one) at any one time. Subsequent routine tests have to be repeated on each sublayer in order to ensure its homogeneity.

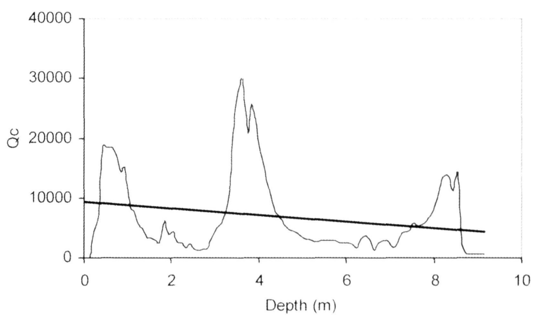
Observing the results of analysis, we could reasonably presumed that these statistical tools are likely to well perform in identifying boundaries at which each divided layer is constituted of linear trend. Nonetheless, the presumption on this limitation does not restrict us from combining two or more layers in the subsequent analyses as far as they possess very similar variation characteristics, i.e. scale of variance and the autocovariance distance (or scale of fluctuation). The modeling which reasonably simplifies the soil profile into fewer layers within the same geological formation and at the same time remains most of the important information is always of great interest from the pragmatic stand.

New Approximate Window Width Estimator

It has been proven through the above case studies that the optimize window width for searching the soil layer boundaries is dependent on the average layer thickness. The lower limit of 1.0 m tends to create plenty of unnecessary noises which might complicate the interpretation. The resistance zone of 1.5 m appears to be too restrictive and may only applicable in certain specific profile. The agreement met in previous works using 1.0 m and 1.5 m window widths are likely to be coincident. The average distance between boundaries from autocorrelation analysis seems to be most relevant, flexible and useful, however relatively complex and time-consuming. In view of these shortcomings, a simple approximate method is proposed to estimate a suitable window width, which is close enough to the expected distance between boundaries that has been justified above, for initial identification of soil layer boundaries.

In random field characterization (Vanmarcke, 1977), autocorrelation analysis has been used to compute the scale of fluctuation – the distance within which soil property shows relatively strong correlation from point to point. The scale of fluctuation can also be defined as a distance over which the soil property is likely to stay either above or below its average value. Also, it is said to be closely related to the average distance between intersections of the fluctuating property and its mean. This is essentially the average distance between boundaries as suggested by Webster (1973). Keeping this in mind, the approximate estimate of scale of fluctuation using the average distance between 'mean-crossings' as advocated by Vanmarcke (1977) could be a simple and feasible alternative to be exploited in making the first approximation on suitable window width.

In light of the above relationship, the cone tip resistance profiles have been fitted with a simple linear trend function using ordinary least square (OLS) regression analysis and the average distance between 'mean-crossings' was evaluated. Figure 5 illustrates the computation of the expected distance for the typical cone tip resistance profiles for the ILNWULAK site. Exercising some cautious evaluations, the estimated average distances between 'mean-crossings' generally fall within the ranges of the suitable window width determined from autocorrelation analysis (Table 1). As a result, the simple approximation of the expected distances between 'mean-crossings' could be inferred as the appropriate window width.



Depth of crossing	Distance between crossing
0.34	
1.04	0.7
3.24	2.2
4.44	1.2
7.54	3.1
8.64	1.1

Average = 1.66 m

Figure 5: Estimation of Expected Distance between 'Mean-crossings' for ILNWULAK Site

Table 1: Summary of Suitable Window Width for Identification of Soil Layer Boundaries

Case study	Autocorrelation analysis		Expected distance between 'mean-crossings'
	2/3 of expected distance	Full expected distance	
CATIFS site	1.9 m	2.8 m	2.1 m
MAUMASSA site	2.4 m	3.6 m	2.7 m
ILNWULAK site	1.3 m	1.9 m	1.7 m

One of the difficulties as mentioned by Webster (1973) and Zhang and Tumay (1996) was solving the profile with layer thicknesses differ widely. To reduce the complication, the worker must first be clear on the object and properly plan before one starts the analysis. For instance, if in the first place, the rough approximate average thickness between layers (by visual) is about 3.0 m, then the optimize window width would probably be around that value or somewhat slightly smaller. Any attempt that far too small or far too large from that value would be fruitless. Often, one time demarcation might not be adequate; on the other extent, excessive subdivision and modeling which has no practical value should be avoided.

In the case whereby a homogeneous layer is of clear evident, for instance the clay layer from approximately 12.0 m depth to the end of exploration of about 30.0 m at the CATIFS site, that section of profile should not be mixed together with other relatively thinner layers at upper depth in the analysis (note that the clay layer from 12.0 m to 30.0 m had been excluded in the first case study here). Or else, any possible erroneous peaks appear within that section should be discarded after incorporated with visual observation and engineering judgment. For verification, the demarcated sections should be examined using stationarity tests, e.g. Kendall's τ test, run test, sign test, etc. which are not covered in this paper. Every method has its own limitations and the underlying concepts must be well understood in order to fully exploit it appropriately.

Conclusion

Generally, both statistical tools can be utilized in identifying soil layer boundaries. RI appears to be more powerful as it can capture most of the prominent major boundaries at one time fairly accurately. Also, it is relatively more robust and persistent able to detect the main peaks at the same positions even with fairly different window widths. The suitable window width used in conjunction with RI was proven as a function of average distance between boundaries which could be determined from autocorrelation analysis. The empirical criterion of 0.7 was found useful in guiding whether a peak is significant enough to be considered as a valid boundary. On the other hand, Bstat was found relatively less useful as no applicable rejection criterion can be referred to in judging the existence or nonexistence of a boundary. In addition, Bstat can only identify limited layer boundaries (usually those most prominent one) at any one time. Subsequent routine tests have to be repeated on each sublayer in order to ensure its homogeneity.

The lower limit of 1.0 m, that tends to create plenty of unnecessary peaks that imply soil boundaries, which might complicate the interpretation. The resistance zone of 1.5 m appears to be too restrictive and may only be applicable in certain specific profiles. The average distance between boundaries from autocorrelation analysis seems to be most relevant, flexible and useful, however relatively complex and time-consuming. In this paper, a simple approximate method is proposed to estimate a suitable window width using the concept of average distance between 'mean-crossings'. The approach was exploited and substantiated as a simple, quick and accurate estimator in making the first approximation on suitable window width for boundary identification exercise.

References

Baecher, G. B., and Christian, J. T. (2003). *Reliability and statistics in geotechnical engineering.* Wiley, Chichester, U.K.

Brockwell, P. J., and Davis, R. A. (2002). *Introduction to time series and forecasting.* 2nd Ed., New York: Springer-Verlag.

Cafaro, F., and Cherubini, C. (2002). "Large sample spacing in evaluation of vertical strength variability of clayey soil." *J. Geotech. Geoenviron. Eng.*, 128(7), 558-568.

Casagrande, A. (1965). "Role of the 'calculated risk' in earthwork and foundation engineering." *J. Soil Mech. Found. Div., Am. Soc. Civ. Eng.*, 91(SM4), 1-40.

Christian, J. T. (2004). "Geotechnical engineering reliability: How well do we know what we are doing?" *J. Geotech. Geoenviron. Eng.*, 130(10), 985-1003.

DeGroot, D. J., and Baecher, G. B. (1993). "Estimating autocovariance of in-situ soil properties." *J. Geotech. Eng.*, 119(1), 147-166.

Fenton, G. A. (1999). "Estimation for stochastic soil models". *J. Geotech. Geoenviron. Eng.,* 125(6), 470-485.

Hegazy, Y. A., Mayne, P. W., and Rouhani, S. (1996). "Geostatistical assessment of spatial variability in piezocone tests". *Uncertainty in the geologic environment: From theory to practice (GSP 58)*, ASCE, New York, 254-268.

Jaksa, M. B., Brooker, P. I., and Kaggwa, W. S. (1997). "Inaccuracies associated with estimating random measurement errors". *J. Geotech. Geoenviron. Eng.,* 123(5), 393-401.

Kulatilake, P. H. S. W., and Um, J. (2003). "Spatial variation of cone tip resistance for the clay site at Texas A&M University". *Probabilistic Site Characterization at the National Geotechnical Experimentation Sites (GSP 121)*, ASCE, New York, 41-60.

Leonards, G. A. (1982). "Investigation of failures". *J. Geotech. Eng.*, 108(GT2), 185-246.

Morgenstern, N. R. (1995). "Managing risk in geotechnical engineering". The 3rd Casagrande Lecture. *Proc., 10th Pan-American Conference on Soil Mechanics and Foundation Engineering*, Guadalajara, 4, 102-126.

National Cooperative Highway Research Program (NCHRP). (2007). *NCHRP Synthesis 368: Cone Penetration Testing – A Synthesis of Highway Practice.* Transportation Research Board of the National Academies, Washington, D. C.

Peck, R. B. (1969). "Advantages and limitations of the observational method in applied soil mechanics". *Geotechnique*, 19(2), 171-187.

Phoon, K. K., and Kulhawy, F. H. (1999). "Characterization of geotechnical variability". *Can. Geotech. J.*, 36, 612-624.

Phoon, K. K., Quek, S. T., and An, P. (2003). "Identification of statistically homogeneous soil layers using modified Bartlett statistics". *J. Geotech. Geoenviron. Eng.*, 129(7), 649-659.

Robertson, P. K. (1986). "In-situ testing and its application to foundation engineering". *Can. Geotech. J.*, 23(3), 573-587.

Tang, W. H. (1984). "Principles of probabilistic characterizations of soil properties". *Probabilistic characterization of soil properties: Bridge between theory and practice*, D. S. David and H.-Y. Ko, eds., ASCE, New York, 74-89.

Vanmarcke, E. H. (1977). "Probabilistic modeling of soil profiles". *J. Geotech. Eng. Div., Am. Soc. Civ. Eng.*, 103(GT11), 1227-1246.

Webster, R. (1973). "Automatic soil boundary location from transect data". *J. Math. Geol.*, 5(1), 27-37.

Whitman, R. V. (2000). "Organizing and evaluating uncertainty in geotechnical engineering". *J. Geotech. Geoenviron. Eng.*, 126(7), 583-593.

Wickremesinghe, D., and Campanella, R. G. (1991). "Statistical methods for soil layer boundary location using the cone penetration test". *Proc., ICASP6*, Mexico City, 636-643.

Wickremesinghe, D. S. (1989). "Statistical characterization of soil profile using in-situ tests". PhD thesis, Univ. of British Columbia, Vancouver, Canada.

Wu, T. H. (1974). "Uncertainty, safety and decision in soil engineering". *J. Geotech. Eng. Div., Am. Soc. Civ. Eng.*, 100(GT3), 329-348.

Zhang, Z. J., and Tumay, M. T. (1996). "The reliability of soil classification derived from cone penetration test". *Uncertainty in the geologic environment: From theory to practice (GSP 58)*, ASCE, New York, 383-408.